Consistency of kernel variance estimators for sums of semiparametric linear processes

James Davidson
Cardiff University

Robert M. de Jong
Michigan State University

This version, January 2002

Abstract

Conditions are derived for the consistency of kernel estimators of the variance of a sum of dependent heterogeneous random variables, with a representation as moving averages of near-epoch dependent functions of a mixing process. Fourth moments are not gnerally required. The conditions permit more dependence than a purely nonparametric representation allows, and may be close to those of the best-known conditions for the functional central limit theorem. The class of permitted kernel functions is different from those usually considered, but can approximate most of the usual choices arbitrarily closely, and can be extended to include them subject to a seemingly innocuous extra condition on the random process.

1 Introduction

Suppose that $\{X_t, t = 1, ..., n\}$ is a sequence of zero-mean random variables, and $s_n^2 = E(\sum_{t=1}^n X_t)^2$, and a functional central limit theorem (FCLT) holds for these data, such that if

$$X_n(r) = s_n^{-1} \sum_{t=1}^{[nr]} X_t \quad 0 \le r \le 1 \tag{1.1}$$

then $X_n \xrightarrow{d} B$, where '$\xrightarrow{d}$' denotes convergence in distribution and B is Brownian motion on the unit interval. Of course, the special case of $X_n(1) \xrightarrow{d} B(1)$ is the regular central limit theorem (CLT). Recent work by the authors (De Jong 1997, De Jong and Davidson 2000b) has studied sufficient conditions for weak convergence in a nonparametric setting, adopting the concept of near-epoch dependence on a mixing process to constrain dependence. Davidson (2002) considers how these conditions may be established in the context of nonlinear time series processes. He also considers semiparametric linear processes, whose driving processes may exhibit unspecified nonlinear dependence, and gives conditions for that case which appear close to necessary.

To make use of these results for asymptotic statistical inference, however, we must possess a consistent estimator $\hat{s}_n$ of s_n. Without a parametric setup, kernel estimators are the natural vehicle for this type of analysis. For example, they are sufficient to compute the Phillips-Perron (1988) test of the unit root hypothesis, the Kwiatkowski, Phillips, Schmidt and Shin (1992) test of the I(0) hypothesis, the Phillips-Ouliaris (1990) nonparametric tests for cointegration, the fully modified least squares estimator of Phillips and Hansen (1990), and related procedures.

Therefore, a key question in this type of analysis is whether sufficient conditions for the various modes of weak convergence are also sufficient for the consistency of the variance estimator. Among studies that have analysed the consistency of kernel estimators are Newey and West (1987), Gallant and White (1988), Andrews (1991), Pötscher and Prucha (1991), Andrews and Monahan (1992), Hansen (1992), and De Jong and Davidson (2000a). However, all these studies impose conditions stronger than the best-known conditions for the application of a FCLT or CLT to the same variables. These issues are discussed in De Jong and Davidson (2000a), where it is noted that the previous studies impose either a form of stationarity, or uniform boundedness in L_p-norms for some $p \ge 2$, precluding the possibility of trending moments. All except Pötscher and Prucha (1991) assume that the random variables considered are strong or uniform mixing and that the true variance converges to some well-defined limit. All except Hansen (1992) assume that the random variables under consideration possess finite fourth moments.[1] All these conditions can be relaxed for the proof of the CLT (see Davidson 1992, 1993 and De Jong 1997), and many of them also for the FCLT (De Jong and Davidson 2000b).

De Jong and Davidson (2000a) presents a consistency theorem under conditions similar to those of the latter weak convergence results, which are the best such conditions currently known to us when the representation of the dependence is purely nonparametic. That result is given for near-epoch dependent functions of mixing processes, and for the class of kernel estimators for which the kernel function possesses a positive Fourier transform, equivalent to the class that necessarily generate positive semidefinite covariance matrix estimates. In this paper, we establish a comparable result for the semiparametric linear processes of the type analysed in Davidson (2002). The processes are moving averages with absolutely summable coefficients of sequences of the general type dealt with in our earlier paper. Linear processes driven by underlying shock sequences with a nonlinear and possibly unknown dependence structure (GARCH, threshold effects, and the like) are often encountered in the applied literature, and Davidson (2002) shows how such processes may satisfy our present assumptions. Applications apart, however, an important motivation for our work is to demonstrate how the amount of dependence permitted to the series may be extended, critically, by restricting the type of dependence.

The assumptions and main theorem are stated in Section 2, and we give an extension to processes depending on estimated parameters in Section 3, applying results from our previous work. Our basic assumptions are pitched at quite a high level, and we discuss and illustrate these in Sections 4 and 5. We show here that the leading cases are closely comparable to those of the existing literature. Our motivation is to pinpoint the restrictions that bind, and so facilitate relaxing these in future work. The proof of the theorem and associated lemmas is given in Section 6. This proof is completely different from that of our previous result and in principle is more straightforward, being based on a blocking argument and an existing law of large numbers given in De Jong (1997).

We adopt a class of kernel estimators that has not been explicitly considered in previous work. In fact, our basic result specifically applies to cases exhibiting discontinuities, such as the uniform kernel, which most of the above-cited work has ruled out. However, subject to an extra condition that appears innocuous although is difficult to establish in a wholly general context, our kernel class can be extended to include most of the popular cases.

Our analysis is carried out for the case of scalar X_t. Whilst many applications of this theory are multivariate, we note that results for the scalar case might be applied to $X_t = a'Y_t$ where Y_t and a are a random vector and an arbitrary constant vector of the same order. As pointed out by Newey and West (1987), the extension of our results to the multivariate case is therefore very straightforward, and we do not consider it explicitly.

2 Main Result

The estimators to be considered have the generic form

$$\hat{s}_n^2 = \sum_{t=1}^{n} X_t^2 + 2\sum_{m=1}^{\gamma_n} k(m/\gamma_n) \sum_{t=1}^{n-m} X_t X_{t+m}. \tag{2.1}$$

where $k(x)$ is the kernel function. Thus the products $X_t X_{t+m}$ for $m > \gamma_n$ (the bandwidth) are excluded from the estimator. [2]

The time series is assumed to take the following form.

Assumption 2.1 (a) $X_t = \sum_{j=0}^{\infty} \theta_j U_{t-j}$ where $\sum_{j=0}^{\infty} |\theta_j| < \infty$.

(b) $E(U_t) = 0$, and there exists a positive constant sequence c_t such that $\{U_t/c_t\}$ is L_r-bounded, $r \geq 2$, uniformly in t.

[1] See de Jong (2000) for a correction of the Hansen (1992) result.

[2] Using this notation, it is not necessary to consider the case $x < 0$.

(c) U_t is L_2-near epoch dependent[3] of size $-\frac{1}{2}$ with respect to constants d_t on either an α-mixing sequence of size $-r/(r-2)$, with $r > 2$, or a ϕ-mixing sequence of size $-r/2(r-1))$, with d_t/c_t bounded uniformly in t. Under ϕ-mixing, $r = 2$ is permitted if $\{U_t/c_t\}$ is uniformly integrable.

(d) If γ_n is the bandwidth sequence specified in (2.1) then

$$\sup_{-\infty \le i \le [n/\gamma_n]+1} M_{ni} = o(\gamma_n^{-1/2}), \tag{2.2}$$

$$\sup_{h \ge 0} \sum_{i=1}^{[n/\gamma_n]+1} M_{n,i-h}^2 = O(\gamma_n^{-1}), \tag{2.3}$$

where

$$M_{ni} = \max_{(i-1)\gamma_n+1 \le t \le i\gamma_n} c_t/s_n. \tag{2.4}$$

The restrictions relating to the kernel function are as follows.

Assumption 2.2 Defining the class of kernels $\mathcal{T} = \{k : [0,1] \mapsto \mathsf{R}$, left continuous, with $k(0) = 1$, and continuous at 0 and all but a finite number of points $\}$, let one of the following conditions hold:

(i) For some $M < \infty$, the kernel belongs to the class

$$\mathcal{T}_M = \{k_M : k_M(x) = k(j/M),\ j/M \le x < (j+1)/M,\ j = 0, ..., M-1,\ k \in \mathcal{T}\}. \tag{2.5}$$

(ii) The kernel is $k \in \mathcal{T}$ and

$$\frac{1}{s_n^2 \max_{1 \le m \le \gamma_n} |a_{nm}|} \sum_{m=1}^{\gamma_n} a_{nm} \sum_{t=1}^{n-m} X_t X_{t+m} = O_p(1) \tag{2.6}$$

as $n \to \infty$, where

$$a_{nm} = k_M(m/\gamma_n) - k(m/\gamma_n)$$

for some $k_M \in \mathcal{T}_M$ defined by (2.5), and $M < \infty$.

The left continuity is just for convenience here, and could be replaced by right continuity by re-working the argument.

Finally, the bandwidth must satisfy at least the following minimal condition.

Assumption 2.3 $\gamma_n \to \infty$ and $\gamma_n/n \to 0$, as $n \to \infty$.

We will call kernels of the class $\mathcal{T}$ truncated. The member of the class usually called the 'truncated' kernel, having $k(x) = 1$ for $0 \le x < 1$ (and hence also belonging to $\mathcal{T}_M$) should be called the uniform kernel, in this context. Given the truncation requirement, $\mathcal{T}$ is as large a class of kernel functions as can reasonably be specified. Of the popular choices listed by Andrews (1991), the uniform, Bartlett, Parzen and Tukey-Hanning kernels all belong to $\mathcal{T}$, although the Quadratic Spectral kernel does not. All these cases except the uniform kernel (which is not guaranteed to yield a positive estimate) are also covered by the results of De Jong and Davidson (2000a).

Letting $\xrightarrow{pr}$ denote convergence in probability, our result can now be stated as follows.

Theorem 2.1 If Assumptions 2.1, 2.2 and 2.3 hold then

$$\frac{\hat{s}_n^2}{s_n^2} \xrightarrow{pr} 1. \tag{2.7}$$

This result establishes, for example, the asymptotic distribution of the computable quantities $\hat{s}_n^{-1} \sum_{t=1}^{[nr]} X_t$, given that the FCLT holds. Note that for this purpose, it is not necessary to show that $\hat{s}_n^2/n - s_n^2/n \xrightarrow{pr} 0$, nor need s_n^2/n converge, nor be uniformly bounded away from either infinity or zero. This is important, since such behaviour is not incompatible with the CLT, at least.[4]

[3] L_p-near epoch dependence is the property $\|X_t - E_{t-m}^{t+m} X_t\|_p \le d_t \nu_m$ where d_t is a sequence of positive constants and $\nu_m \to 0$. See for example Davidson (1994) for details.

[4] It is incompatible with the usual FCLT, in the sense of weak convergence to B. See Davidson (1994) Section 29.4 for details

3 Estimated Parameters

It is desirable to extend the results to allow X_t to depend on estimated parameters, and allow the bandwidth to depend on a sample-dependent scale factor. However, a number of results of this type exist, such as (respectively) Theorems 2.2 and 2.3 of De Jong and Davidson (2000a). Rather than attempt new proofs under the different kernel assumptions of the present paper, we point out that for the class of kernels $\mathcal{T} \cap \mathcal{K}$ where the class $\mathcal{K}$ is defined in Assumption 1 of De Jong and Davidson (2000a), the two results cited extend to the present case, under our Assumption 2.2(ii), together with Assumption 4 of the cited paper. Of the cases listed following Assumption 2.3, only the uniform kernel is excluded by these assumptions, which in any case is not a recommended choice for the reason mentioned. Be careful to note however that the subclass $\mathcal{T}_M$ of Assumption 2.2(i) is ruled out, because the kernels in class $\mathcal{K}$ are continuous.

We give these facts a formal statement, as follows. Let $X_t(\theta)$ represent a process depending on an unknown parameter with true value θ_0, let $\hat{\theta}_n$ denote a consistent estimator of θ_0, let the bandwidth be chosen as $\hat{\alpha}_n\gamma_n$ where $\hat{\alpha}_n$ is a sample-based stochastic sequence, and let the kernel estimator depending on these quantities be denoted

$$\hat{s}_n^2(\hat{\theta}_n, \hat{\alpha}_n) = \sum_{t=1}^{n} X_t(\hat{\theta}_n)^2 + 2\sum_{m=1}^{[\hat{\alpha}_n\gamma_n]} k(m/(\hat{\alpha}_n\gamma_n)) \sum_{t=1}^{n-m} X_t(\hat{\theta}_n)X_{t+m}(\hat{\theta}_n). \tag{3.1}$$

Theorem 3.1 If in (3.1)

(a) the kernel function satisfies Assumption 2.2, and also Assumption 1 and conditions (2.11) and (2.12) of De Jong and Davidson (2000a);

(b) γ_n satisfies Assumption 2.3, $\hat{\alpha}_n = O_p(1)$, and $\hat{\alpha}_n^{-1} = O_p(1)$;

(c) $X_t(\theta_0)$ satisfies Assumption 2.1 with respect to γ_n;

(d) $X_t(\theta)$ and $\hat{\theta}_n$ satisfy Assumption 4 of Davidson and de Jong (2000a);

then

$$\frac{\hat{s}_n^2(\hat{\theta}_n, \hat{\alpha}_n)}{s_n^2} \xrightarrow{pr} 1.$$

For the convenience of readers, the cited assumptions from De Jong and Davidson (2000a) are summarised in the Appendix. The proof of Theorem 3.1 is by applying Theorems 2.2 and 2.3 of the same paper. Note that the proofs of these latter results depend on assumptions other than Assumption 4 only to establish the consistency of the estimator corresponding to $\hat{s}_n^2$. In the present context, this result can be replaced by Theorem 2.1, applied pointwise for the case of bandwidths $\alpha\gamma_n$ where $\varepsilon < \alpha < 1/\varepsilon$ for $\varepsilon > 0$.

4 Discussion of Assumption 2.1

The conditions in Assumption 2.1 are similar to those given for the FCLT in Davidson (2002), Theorem 3.1. In the latter result the restrictions on the heterogeneity of the sequence are stronger but the restriction on dependence somewhat milder. Absolute summability of the moving average coefficients is not required, although the actual restriction is stronger than square summability. It is worth noting that the absolute summability condition is needed here to validate an application of the triangle inequality (see the proof of Lemma 6.2 below) which does not appear sharp. It is an interesting conjecture that the same conditions suffice for each result, especially since those for the FCLT appear close to necessary. We have not been successful in showing this as a general result, but it should not be too difficult to validate examples such as those discussed in Davidson (2002), Section 3, in which the sign of θ_j switches periodically. This would involve taking the terms of the moving average in blocks whose sum converges absolutely.

Assumption 2.1 allows variances with either positive or negative trends, the case $EU_t^2 = t^\alpha$ being covered for any $\alpha > -1$, or in other words, such that s_n^2 increases. By contrast, Assumptions 2 and 3 of de Jong and Davidson (2000a) allow any rate of increasing trend, but not negative trend. In these cases the FCLT does not hold, in the sense that the limit process is not Brownian motion (see for example Davidson 1994, Section 29.4) although the CLT does hold. Assumption 2.1(d) may then imply an additional restriction on the bandwidth, as shown in Davidson (1993). If the sequence c_t is uniformly bounded away from both 0 and ∞, on the other hand, then equations (2.2) and (2.3) hold for any bandwidth satisfying Assumption 2.3. In practice, considerations such as reducing the mean squared error of estimate (assuming fourth moments of the data exist) may dictate particular choices of bandwidth. See Andrews (1991) and Newey and West (1994) for details of these procedures.

A difference between these assumptions and their counterparts in De Jong (1997) is that the lower limit of the supremum in (2.2) is extended from 1 to $-\infty$, with a related modification to (2.3). Only the case $h = 0$ is considered in De Jong (1997). These extensions are needed to allow for models explicitly involving infinite lags, as in Assumption 2.1(a). Note that if the moving average is actually truncated, with $\theta_j = 0$ either for $j > 0$ (the case $X_t = U_t$), or beyond some finite lag, then the U_t process can be set to 0 prior to the truncation point. Therefore, for such processes, (2.2) and (2.3) can hold trivially and impose no extra restrictions on the heterogeneity.

5 Discussion of Assumption 2.2

The complications in Assumption 2.2, requiring the possible replacement of k by k_M, are essentially technical in character. The approximation of any chosen k by k_M can be arbitrarily close, although note that M must be fixed and may not go to ∞ with n, and if part (ii) of the assumption does not hold, it is not permissible to assume that substituting k for k_M will yield the same limit.

However, it appears a plausible conjecture that Assumption 2.2(ii) holds generally, as a consequence of Assumption 2.1; at any rate, we have not succeeded thus far in constructing a counter-example. The necessary condition that $s_n^{-2}\sum_{t=1}^{n-m} X_t X_{t+m} \xrightarrow{pr} 0$, for all but a finite number of m, is certainly an implication of Assumption 2.1. What is harder to establish by direct means, although intuitively highly plausible, is that the sum of these γ_n covariance terms, with arbitrary weights from the interval $[-1, 1]$ as in (2.6), is $O_p(1)$. Note that the absolute sum of these terms would generally diverge.

To take a related and familiar example, consider $\gamma_n = n$ and $a_{nm} = 1$. After re-ordering, the double sum reduces to

$$s_n^{-2}\sum_{t=1}^{n}\sum_{s=1}^{t-1} X_s X_t \xrightarrow{d} \int_0^1 X dX \tag{5.1}$$

where X is an a.s. continuous Gaussian process, which follows from the assumptions by de Jong and Davidson (2000b), Theorem 4.l. A natural approach to establishing Assumption 2.2(ii) might therefore be to show that (2.6) has a weak limit comparable to (5.1). However, the inclusion of the weights a_{nm} in the sum requires a non-trivial extension of the existing results for stochastic integrals, so for the moment this remains a conjecture.

The direct approach of bounding the second moment can also be applied, in those cases where the calculations are feasible. One way, that may be regarded as adequately general for most applications, is to make the shock process independent. The following result also calls for fourth moments, and while this assumption is pretty standard in the literature, it is one that is avoided in De Jong and Davidson (2000a), for example, and is not necessary for the FCLT.

Theorem 5.1 Let Assumptions 2.1 and 2.3 hold, and let U_t be independent and identically distributed with finite fourth moment. Then Assumption 2.2(ii) holds.

Proof. Let $\sigma^2 = E(U_t^2)$ and $\mu_4 = E(U_t^4)$. Squaring (2.6) and taking expectations, note that $s_n = O(n^{1/2})$ under stationarity, and that the terms have expectation zero unless the random arguments are equal in pairs. Therefore,

$$\begin{aligned} &E\left(s_n^2\sum_{m=1}^{\gamma_n} a_{nm}\sum_{t=1}^{n-m} X_t X_{t+m}\right)^2 \\ &= s_n^{-4}E\left(\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\theta_j\theta_k\sum_{m=1}^{\gamma_n} a_{nm}\sum_{t=1}^{n-m} U_{t-j}U_{t+m-k}\right)^2 \\ &= s_n^{-4}(\mu_4-\sigma^4)n\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\theta_j^2\theta_k^2 a_{n,k-j}^2 I(1 \le k-j \le \gamma_n) \end{aligned}$$

$$
\begin{aligned}
&+ s_n^{-4}\sigma^4 \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\sum_{j'=0}^{\infty}\sum_{k'=0}^{\infty} \theta_j\theta_k\theta_{j'}\theta_{k'} \\
&\times \Bigg(a_{n,k-j}a_{n,k'-j'}(n-k+j)(n-j'+k')I(1 \le k-j \le \gamma_n)I(1 \le k'-j' \le \gamma_n) \\
&+ \sum_{m=1}^{\gamma_n}(n-m)a_{nm}a_{n,m+j-j'-k+k'}I(1 \le m+j-j'-k+k' \le \gamma_n) \\
&+ \sum_{m=1}^{\gamma_n}(n-m)a_{nm}a_{n,k'+k-j'-j-m}I(1 \le k'+k-j'-j-m \le \gamma_n)\Bigg) \\
&= O\big(\max_{1\le m\le \gamma_n} a_{nm}^2\big) \qquad (5.2)
\end{aligned}
$$

noting that, for fixed n, the quadruple sum in (5.2) is bounded by a multiple of $(\sum_{j=0}^{\infty}\theta_j)^4 < \infty$. ∎

To generalize this argument to cases where the random sequence is dependent and nonstationary is possible in principle, but requires showing the summability of sequences of fourth cross-moments of the form $E(U_sU_tU_uU_v)$. It would call for an extension of the current techniques for near-epoch dependent processes, and additional specialized assumptions. On the other hand, if a weak convergence result could be obtained, the example of (5.1) shows how the assumption of finite fourth moments could probably be dispensed with. Note that the number of terms depending on μ_4 in (5.2) is of lower order than those depending on σ^4, so that a normalization exists under which the limiting variance is positive yet does not depend on μ_4. As a familiar parallel, consider the fact that a $\chi^2(1)$ random variable possesses a variance, notwithstanding that it has a representation as the limit of a random sequence of the form $n^{-1}(\sum_{t=1}^{n}X_t)^2$, $n \ge 1$, where $E(X_t^4) = \infty$ is permitted. The variance exists in the limit, but not for any finite n, in this case.

6 Proof of Theorem 2.1

Our approach involves breaking the sum of the X_t variables into blocks of γ_n terms. Let

$$
Z_{ni} = \sum_{t=(i-1)\gamma_n+1}^{i\gamma_n} X_t,\ \ i = 1, \dots, [n/\gamma_n] \qquad (6.1)
$$

and

$$
Z_{n,[n/\gamma_n]+1} = \sum_{t=[n/\gamma_n]\gamma_n+1}^{n} X_t. \qquad (6.2)
$$

The fundamental law of large numbers underlying the result is the following.

Lemma 6.1 If Z_{ni} is defined by (6.1) and (6.2), then under Assumptions 2.1 and 2.3,

$$
\sum_{i=1}^{[n/\gamma_n]+1}(Z_{ni}^2 - EZ_{ni}^2) \xrightarrow{pr} 0.
$$

This is a basic step in the proof of the CLT and FCLT, and hence is to be expected to hold under similar conditions.

The second fundamental requirement, of showing that the truncation is compatible with the consistency of the estimator since the additional terms are negligible, is established as follows.

Lemma 6.2 Under Assumptions 2.1 and 2.3,

$$
\sum_{t=1}^{n}\sum_{m=0}^{n-t}|E(X_tX_{t+m})| = O\Bigg(\sup_{r>0}\sum_{t=1}^{n}c_{t-r}^2\Bigg) \qquad (6.3)
$$

and

$$\sum_{t=1}^{n}\sum_{m=0}^{n-t}|E(X_tX_{t+m})|I(m \geq \gamma_n) = o\left(\sup_{r>0}\sum_{t=1}^{n}c_{t-r}^2\right). \tag{6.4}$$

Now, define

$$A_n = Z_{n1}^2 + \sum_{i=2}^{[n/\gamma_n]+1}(Z_{ni}^2 + 2Z_{n,i-1}Z_{ni}). \tag{6.5}$$

This can be shown to be a form of truncated kernel estimator of s_n^2.[5] The steps are first to show that it is consistent, and then to show the difference between this estimator and members of the general truncated class of Assumption 2.2, suitably normalized, are converging in probability to a limit arbitrarily close to 0. These requirements are met by the next three lemmas, from which Theorem 2.1 follows directly. Under Assumptions 2.1, 2.3 and 2.2:

Lemma 6.3 $s_n^{-2}(A_n - E(A_n)) \xrightarrow{pr} 0.$

Lemma 6.4 $s_n^{-2}E(A_n) - 1 \longrightarrow 0.$

Lemma 6.5 $s_n^{-2}(\hat{s}_n^2 - A_n) \xrightarrow{pr} 0.$

The proofs of the five lemmas are given as follows.

Proof of Lemma 6.1

Without loss of generality, let n increase through a sequence of values such that $r_n = n/\gamma_n$ is always an integer, so that the final term in (6.2) can be ignored. For $i = 1, ..., r_n$, substitute from Assumption 2.1(a) into (6.1) and then re-order the sum, to obtain

$$\begin{aligned} Z_{ni} &= \sum_{t=(i-1)\gamma_n+1}^{i\gamma_n}\sum_{j=0}^{\infty}\theta_jU_{t-j} \\ &= \sum_{t=(i-1)\gamma_n+1}^{i\gamma_n}\psi_{1nt}U_t + \sum_{t=-\infty}^{(i-1)\gamma_n}\psi_{2nt}U_t \\ &= Z_{1ni} + Z_{2ni} \end{aligned} \tag{6.6}$$

(defining Z_{1ni} and Z_{2ni}) where

$$\psi_{1nt} = \sum_{j=0}^{i\gamma_n - t}\theta_j \qquad (i-1)\gamma_n + 1 \leq t \leq i\gamma_n \tag{6.7}$$

$$\psi_{2nt} = \sum_{j=(i-1)\gamma_n+1-t}^{i\gamma_n - t}\theta_j \qquad \infty < t \leq (i-1)\gamma_n \tag{6.8}$$

Observe that the arrays ψ_{1nt} and ψ_{2nt} do not depend on i except through t. Their patterns within each block are the same, in the first case comprising the sums of the first m MA coefficients, for $m = i\gamma_n - t$, in descending order from γ_n to 1, and in the second case the sums of γ_n consecutive MA coefficients starting at the mth, for $m = (i-1)\gamma_n - t \geq 1$.

Next note that

$$\|Z_{2ni}\|_2 \leq s_n \sup_{j<i} M_{nj}K_n \tag{6.9}$$

[5] The best way to visualize the construction is by considering the square array of the n^2 cross-product terms. A_n contains the terms X_tX_{t+m} lying in a "saw-tooth" band about the main diagonal, such that $-K\gamma_n \leq m \leq K\gamma_n$ for values of $K \in [1, 2]$.

by Assumption 2.1(b), where

$$K_n^2 = \sum_{k=0}^{\infty}\left(\sum_{j=1+k}^{\gamma_n+k}\theta_j\right)^2 = o(\gamma_n)$$

holds as a consequence of absolute summability, see Davidson (2002) Section 3. Now,

$$\sum_{i=1}^{r_n}(Z_{ni}^2 - EZ_{ni}^2) = \sum_{i=1}^{r_n}(Z_{1ni}^2 - EZ_{1ni}^2) + \sum_{i=1}^{r_n}(Z_{2ni}^2 - EZ_{2ni}^2)$$
$$+ 2\sum_{i=1}^{r_n}(Z_{1ni}Z_{2ni} - EZ_{1ni}Z_{2ni}). \tag{6.10}$$

We show the convergence in probability of each of the three right-hand side terms. First, note that $\psi_{1nt} = O(1)$ as $n \to \infty$ by assumption, and therefore the random variables $X_{nt} = s_n^{-1}\psi_{1nt}U_t$ satisfy Assumption 2 of De Jong (1997). It follows by Lemma 5 of the same source that

$$\frac{1}{s_n^2}\sum_{i=1}^{r_n}(Z_{1ni}^2 - EZ_{1ni}^2) \xrightarrow{pr} 0.$$

Next, applying the Jensen inequality, then Minkowski's inequality and (6.9), note that

$$\left\|\sum_{i=1}^{r_n}(Z_{2ni}^2 - EZ_{2ni}^2)\right\|_1 \leq 2\left\|\sum_{i=1}^{r_n} Z_{2ni}^2\right\|_1$$
$$\leq 2s_n^2K_n^2\sup_{h>0}\sum_{i=1}^{r_n} M_{n,i-h}^2 = o(s_n^2)$$

by Assumption 2.1(d). Finally, in a similar manner but also using the Cauchy-Schwarz inequality,

$$\left\|\sum_{i=1}^{r_n}(Z_{1ni}Z_{2ni} - EZ_{1ni}Z_{2ni})\right\|_1 \leq 2\sum_{i=1}^{r_n}\|Z_{1ni}\|_2\,\|Z_{2ni}\|_2$$
$$\leq 2s_n^2K_n\sup_{h>0}\sum_{i=1}^{r_n} M_{ni}M_{n,i-h}$$
$$\leq 2s_n^2K_n\sup_{h\geq 0}\sum_{i=1}^{r_n} M_{n,i-h}^2 = o(s_n^2).$$

∎

Proof of Lemma 6.2

We show (6.4), but then (6.3) will follow directly from the same arguments, by setting $\gamma_n = 0$ in the formulae.

The triangle inequality gives

$$\sum_{t=1}^{n}\sum_{m=0}^{n-t}|EX_tX_{t+m}|I(m \geq \gamma_n) \leq \sum_{j=0}^{\infty}\sum_{k=0}^{\infty}|\theta_j||\theta_k|\sum_{t=1}^{n}\sum_{m=0}^{n-t}|EU_{t-j}U_{t+m-k}|I(m \geq \gamma_n) \tag{6.11}$$

Letting $r = j-k$ and $v = m+r$, consider without loss of generality the terms $|EU_tU_{t+v}|$. Under Assumption 2.1, U_t is a mixingale sequence of size $-\frac{1}{2}$ with respect to constants c_t, by Davidson (1994) Theorem 17.6. Define $\xi_{tp} = E(U_t|\mathcal{F}_{t-p})$ and $\zeta_{tp} = U_t - \xi_{tp}$, and note that

$$E\xi_{tp}^2 - E\xi_{t,p+1}^2 = E\zeta_{t,p+1}^2 - E\zeta_{tp}^2 \tag{6.12}$$

and that $\xi_{tp} = O_p(c_tp^{-1/2-\delta})$ and $\zeta_{t,-p} = O_p(c_tp^{-1/2-\delta})$ as $p \to \infty$ for $\delta > 0$, by the mixingale property. According to Lemma 3 of De Jong (1997),

$$|EU_tU_{t+v}| \leq \sum_{p=-\infty}^{\infty}\left(E\xi_{tp}^2 - E\xi_{t,p+1}^2\right)^{1/2}\left(E\xi_{t+v,p+v}^2 - E\xi_{t+v,p+v+1}^2\right)^{1/2}. \tag{6.13}$$

Rearrangement of the sums therefore yields

$$\begin{aligned}
&\sum_{t=1}^{n}\sum_{v=r}^{n-t+r}|EU_tU_{t+v}|I(v>\gamma_n)\\
&\qquad\le \sum_{p=-\infty}^{\infty}\sum_{t=1}^{n}\left(E\xi_{tp}^2-E\xi_{t,p+1}^2\right)^{1/2}\sum_{v=r}^{n-t+r}\left(E\xi_{t+v,p+v}^2-E\xi_{t+v,p+v+1}^2\right)^{1/2}I(v>\gamma_n)\\
&\qquad= T_1+T_2+T_3+T_4
\end{aligned}$$

where

$$\begin{aligned}
T_1&=\sum_{p=0}^{\infty}\sum_{t=1}^{n+\min\{0,r\}}\left(E\xi_{tp}^2-E\xi_{t,p+1}^2\right)^{1/2}\sum_{v=\max\{0,r\}}^{n-t+r}\left(E\xi_{t+v,p+v}^2-E\xi_{t+v,p+v+1}^2\right)^{1/2}I(v>\gamma_n)\\
T_2&=\sum_{p=0}^{\infty}\sum_{t=1+r}^{n+r}\left(E\xi_{tp}^2-E\xi_{t,p+1}^2\right)^{1/2}\sum_{v=\max\{1,1-t\}}^{-r}\left(E\xi_{t+v,p+v}^2-E\xi_{t+v,p+v+1}^2\right)^{1/2}I(v>\gamma_n)\\
T_3&=\sum_{p=1}^{\infty}\sum_{t=1+\max\{0,r\}}^{n+r}\left(E\zeta_{t,1-p}^2-E\zeta_{t,-p}^2\right)^{1/2}\\
&\qquad\times\sum_{v=\max\{0,r\}}^{t-1}\left(E\zeta_{t-v,1-(p+v)}^2-E\zeta_{t-v,-(p+v)}^2\right)^{1/2}I(v>\gamma_n)\\
T_4&=\sum_{p=1}^{\infty}\sum_{t=1}^{n}\left(E\zeta_{t,1-p}^2-E\zeta_{t,-p}^2\right)^{1/2}\\
&\qquad\times\sum_{v=\max\{1,t-n-r\}}^{-r}\left(E\zeta_{t+v,1-(p+v)}^2-E\zeta_{t+v,-(p+v)}^2\right)^{1/2}I(v>\gamma_n).
\end{aligned}$$

In these formulae, note how the roles of leads and lags have been interchanged in the pairs T_1, T_2, and T_3, T_4, to ensure that $p+v$ is always nonnegative and increasing in p and v. Applying the convention that empty sums equal zero, note that T_2 and T_4 vanish unless $r<0$, and T_1 and T_3 vanish unless $r>-n$. Using the argument of de Jong (1997) Lemma 4, it can be shown that each of these terms is $O(\gamma_n^{-2\delta}\sup_{r>0}\sum_{t=1}^{n}c_{t-r}^2)$. Combining these results with (6.11) and Assumption 2.1(d) gives

$$\begin{aligned}
\sum_{t=1}^{n}\sum_{m=0}^{t-1}|EX_tX_{t-m}|I(m\ge\gamma_n)&\le C_1\left(\sup_{r>0}\sum_{t=1}^{n}c_{t-r}^2\right)\\
&\times\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\theta_j\theta_k(I(\gamma_n-j+k>0)(\gamma_n-j+k)^{-2\delta}+I(\gamma_n-j+k\le 0))
\end{aligned}\tag{6.14}$$

for $C_1>0$. Since $|\theta_j|=o(j^{-1})$ by Assumption 2.1(a), standard summability arguments (see for example Davidson and de Jong (2000, Lemma A.1)) yield the result that

$$\begin{aligned}
&\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}\theta_j\theta_k(I(\gamma_n-j+k>0)(\gamma_n-j+k)^{-2\delta}+I(\gamma_n-j+k\le 0))\\
&\qquad=\sum_{k=0}^{\infty}\theta_k\left(\sum_{j=0}^{\gamma_n+k}\theta_j(\gamma_n-j+k)^{-2\delta}+\sum_{j=\gamma_n+k+1}^{\infty}\theta_j\right)\\
&\qquad=o(1).
\end{aligned}\tag{6.15}$$

This completes the proof of (6.4). ∎

Proof of Lemma 6.3

We can write

$$\begin{aligned} A_n &= Z_{n1}^2 + \sum_{i=2}^{[n/\gamma_n]+1} (Z_{ni}^{*2} - Z_{n,i-1}^2) \\ &= \sum_{i=1}^{r_{1n}} Z_{n,2i}^{*2} + \sum_{i=1}^{r_{2n}} Z_{n,2i+1}^{*2} - \sum_{i=1}^{[n/\gamma_n]} Z_{ni}^2 \\ &= A_{1n} + A_{2n} - A_{3n} \end{aligned} \tag{6.16}$$

where

$$Z_{ni}^* = Z_{ni} + Z_{n,i-1} = \sum_{t=(i-2)\gamma_n+1}^{i\gamma_n} X_t, \ i = 2, ..., [n/\gamma_n] \tag{6.17}$$

$$Z_{n,[n/\gamma_n]+1}^* = \sum_{t=([n/\gamma_n]-1)\gamma_n+1}^{n} X_t \tag{6.18}$$

$$r_{1n} = \begin{cases} [n/(2\gamma_n)] + 1, & \text{if } [n/\gamma_n] \text{ is odd} \\ [n/(2\gamma_n)], & \text{if } [n/\gamma_n] \text{ is even} \end{cases} \tag{6.19}$$

and $r_{2n} = [n/(2\gamma_n)]$. Writing the identity

$$A_n - E(A_n) = (A_{1n} - E(A_{1n})) + (A_{2n} - E(A_{2n})) - (A_{3n} - E(A_{3n})) \tag{6.20}$$

it is clear that the convergence in probability to zero of the $(A_{in} - E(A_{in}))/s_n^2$, for i =1, 2, and 3, each follows by Lemma 6.1, noting that the order-of-magnitude assumptions on γ_n also hold for $2\gamma_n$. ∎

Proof of Lemma 6.4

$$\begin{aligned} s_n^{-2}|s_n^2 - E(A_n)| &= 2s_n^{-2}\left|\sum_{i=2}^{[n/\gamma_n]+1} \sum_{j=2}^{i-[n/\gamma_n]-1} E(Z_{ni}Z_{n,i-j})\right| \\ &\leq s_n^{-2} \sum_{t=1}^{n} \sum_{m=0}^{n-t} |E(X_tX_{t+m})| I(m \geq \gamma_n) \end{aligned} \tag{6.21}$$

and convergence to zero of the majorant term in equation (6.21) follows by Lemma 6.2. ∎

Proof of Lemma 6.5

For a fixed integer $M > 0$, and a kernel $k \in \mathcal{T}$ that is also continuous everywhere on $[0, 1)$, let $\hat{s}_{nM}^2$ denote the kernel estimator employing the left-continuous kernel k_M defined in (2.5). Write

$$\frac{A_n - \hat{s}_n^2}{s_n^2} = \frac{A_n - \hat{s}_{nM}^2}{s_n^2} + \frac{\hat{s}_{nM}^2 - \hat{s}_n^2}{s_n^2} \tag{6.22}$$

where the second term vanishes under Assumption 2.2(i). Otherwise, note that

$$\frac{\hat{s}_{nM}^2 - \hat{s}_n^2}{s_n^2} = \frac{2}{s_n^2} \sum_{m=1}^{\gamma_n} a_{nm} \sum_{t=1}^{n-m} X_tX_{t+m} \tag{6.23}$$

and

$$\max_{1 \leq m \leq \gamma_n} |a_{nm}| \to \sup_{x \in [0,1)} |k_M(x) - k(x)| \tag{6.24}$$

as $n \to \infty$. Under Assumption 2.2(ii) and the assumed continuity of $k(\cdot)$, (6.23) therefore be can be made as small in probability as desired, by choosing M large enough.

To show that the first right-hand side term converges in probability to 0, first re-order the summation. Write

$$A_n = \sum_{m=0}^{\infty} W_{nm} \tag{6.25}$$

where

$$W_{n0} = \sum_{t=1}^{n} X_t^2 \tag{6.26}$$

$$W_{nm} = 2 \sum_{t=1+m}^{n} X_t X_{t-m} \tag{6.27}$$

for $m = 1, ..., \gamma_n$,

$$W_{nm} = 2 \sum_{i=2}^{[n/\gamma_n]} \left(\sum_{t=(i-2)\gamma_n+m}^{i\gamma_n} X_t X_{t-m} \right) + 2 \sum_{t=([n/\gamma_n]-1)\gamma_n+m}^{n} X_t X_{t-m} \tag{6.28}$$

for $m = \gamma_n + 1, ..., 2\gamma_n$, where the final sum is taken to equal 0 if the lower limit exceeds the upper, and $W_{nm} = 0$ for $m > 2\gamma_n$. Then, note that

$$\frac{A_n - \hat{s}_{nM}^2}{s_n^2} = \sum_{j=1}^{M-1} (1 - k(j/M)) P_{nj} + \frac{1}{s_n^2} \sum_{m=\gamma_n}^{2\gamma_n} W_{nm} \tag{6.29}$$

where

$$P_{nj} = \frac{1}{s_n^2} \sum_{m=[j\gamma_n/M]}^{[(j+1)\gamma_n/M]-1} W_{nm}. \tag{6.30}$$

Note that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{s_n^2} \sum_{m=0}^{2\gamma_n} (W_{nm} - EW_{nm}) = 0 \tag{6.31}$$

by Lemma 6.3, and also that

$$\lim_{K\to\infty} \limsup_{n\to\infty} \frac{1}{s_n^2} \sum_{m=K+1}^{\infty} |EW_{nm}| = 0 \tag{6.32}$$

in view of Lemma 6.2 and the fact that $\sup_{r>0} \sum_{t=1}^{n} c_{t-r}^2 = O\left(s_n^2\right)$ by (2.3). Therefore,

$$\sum_{j=N}^{M-1} P_{nj} + \frac{1}{s_n^2} \sum_{m=\gamma_n}^{2\gamma_n} W_{nm} \xrightarrow{pr} 0 \tag{6.33}$$

for each $1 \leq N < M$ and $M > 0$. Considering (6.33) for the cases $N = 1$ and $N = 2$, it follows by Slutsky's Theorem that $P_{n1} \xrightarrow{pr} 0$, and hence $(1 - k(1/M))P_{n1} \xrightarrow{pr} 0$. Arguing similarly for $N = 2, ..., M - 1$ shows that each of the M right-hand side terms in (6.29) converges in probability to 0, and hence their sum so converges.

The result is extended to kernels with a finite number of left discontinuities by adding these points to the set of jumps defining k_M, so the proof is complete. ∎

Appendix: The Assumptions of De Jong and Davidson (2000a)

The assumptions cited in Theorem 3.1 are as follows. Assumption 1 cited in condition (a) imposes conditions on the kernel function, requiring that $k(.) \in \mathcal{K}$ where

$$\mathcal{K} = \{k(.) : \mathsf{R} \to [-1, 1] \mid k(0) = 1, k(x) = k(-x)\ \forall x \in \mathsf{R},$$

$$\int_{-\infty}^{\infty} |k(x)|dx < \infty, \int_{-\infty}^{\infty} |\psi(\xi)|d\xi < \infty,$$

$$k(.) \text{ is continuous at 0 and at all but a finite number of points}\},$$

where

$$\psi(\xi) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x)e^{i\xi x}dx.$$

To allow stochastic bandwidths, these must be strengthened by the further conditions

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |k(\alpha x)|dx < \infty.$$

and

$$\int_{-\infty}^{\infty} \sup_{\alpha \in [\varepsilon, 1/\varepsilon]} |\psi(\alpha \xi)|d\xi < \infty.$$

which are respectively equations (2.11) and (2.12) of the paper. Note that this full set of conditions is fulfilled by the Bartlett, Parzen, Quadratic Spectral and Tukey-Hanning kernels.

Assumption 4 cited in condition (d) represents alternative smoothness conditions on the stochastic process as a function of unknown parameters. It has the following three parts:

(a) $n^{1/2}\kappa_n(\hat{\theta}_n - \theta_0) = O_p(1)$, where $\kappa_n = \text{diag}(\kappa_{1n}, ..., \kappa_{rn})$ is a deterministic $r \times r$ matrix;

(b) $n^{-1/2}\kappa_n^{-1} \sum_{t=1}^n E\partial X_{nt}(\theta)/\partial\theta$ is continuous at θ_0 uniformly in n;

(c) there exists $\mathcal{N} \subset \Theta$, an open neighbourhood of θ_0, such that,

$$\limsup_{n\to\infty} \sum_{t=1}^n E \sup_{\theta\in\mathcal{N}} \left| \kappa_n^{-1} \frac{\partial X_{nt}(\theta)}{\partial\theta'} \right|^2 < \infty,$$

and either (i)

$$\sup_{\theta\in\mathcal{N}} \left\| n^{-1/2}\kappa_n^{-1} \sum_{t=1}^n e^{i\xi t/\gamma_n} \left(\frac{\partial X_{nt}(\theta)}{\partial\theta'} - E\frac{\partial X_{nt}(\theta)}{\partial\theta'} \right) \right\|_2 \to 0$$

for all $\xi \in \mathsf{R}$, or (ii), $n^{-1/2}\gamma_n = o(1)$ and

$$\sup_{\theta\in\mathcal{N}} \left| \sum_{t=1}^n \kappa_n^{-1} \frac{\partial X'_{nt}(\theta)}{\partial\theta} \frac{\partial X_{nt}(\theta)}{\partial\theta'} \kappa_n^{-1} \right| = O_p(1).$$

See the paper for a full discussion of these conditions, which are rather mild. In the case of linear functions they reduce merely to standard consistency conditions, and are certainly fulfilled under the present Assumption 2.1 applied to the constituent processes. In addition, they allow deterministic and/or stochastic trends, and different rates of convergence of the parameters when $\kappa_n \neq I$.

References

Andrews, D. W. K. (1991) Heteroscedasticity and autocorrelation consistent covariance matrix estimation, Econometrica, 59, 817–858.

Andrews, D. W. K. and J. C. Monahan (1992) An improved heteroscedasticity and autocorrelation consistent covariance matrix estimator, Econometrica 60, 953–966 .

Davidson, J. (1992) A central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes, Econometric Theory 8, 313–334.

Davidson, J. (1993) The central limit theorem for globally nonstationary near-epoch dependent functions of mixing processes: the asymptotically degenerate case, *Econometric Theory* 9, 402–412.

Davidson, J. (1994) *Stochastic Limit Theory*, Oxford: Oxford University Press.

Davidson, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243–269.

Davidson, J. and R. M. De Jong (2000) The functional central limit theorem and convergence to stochastic integrals II: fractionally integrated processes. *Econometric Theory* 16, 643–666.

Gallant, A. R. and H. White (1988) *A unified theory of estimation and inference for nonlinear dynamic models*, New York: Basil Blackwell.

Hansen, B. E. (1992) Consistent covariance matrix estimation for dependent heterogeneous processes, *Econometrica*, 60, 967–972.

De Jong, R. M. (1997) Central limit theorems for dependent heterogeneous processes, *Econometric Theory* 13, 353–367.

De Jong, R. M. (2000) A strong consistency proof for heteroscedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory* 16, 262–267.

De Jong, R. M. and J. Davidson (2000a) Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices, *Econometrica* 68, 407–424.

De Jong R. M. and J. Davidson (2000b) The functional central limit theorem and convergence to stochastic integrals I: weakly dependent processes. *Econometric Theory* 16, 621–642.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics* 54, 159–178.

Newey, W. K. and K. D. West (1987) A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Newey, W. K. and K. D. West (1994): Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, 61, 631–654.

Phillips, P.C.B. and Hansen, B.E. (1990) Statistical inference in instrumental variables regression with I(1) Processes, *Review of Economic Studies* 57, 99–125..

Phillips, P. C. B. and P. Perron (1988) Testing for a unit root in time series regression, *Biometrika* 75, 335–346.

Phillips, P. C. B. and S. Ouliaris (1990), Asymptotic properties of residual based tests for cointegration, *Econometrica* 58, 164–194.

Pötscher, B. M. and I. R. Prucha (1991) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part 2: asymptotic normality, *Econometric Reviews* 10, 253–325